UNIVERSITAT DE BARCELONA

Grado de Lingüística

Trabajo de Fin de Grado

Curso 2017-2018

MINERÍA DE FALACIAS

EN EL DISCURSO POLÍTICO

Gordon Wells

TUTORAS:

María Antònia Martí
Mariona Taulé

Barcelona, 20 de junio de 2018

CONTENIDO

1 INTRODUCCIÓN

En los últimos años hemos visto grandes avances en el área del procesamiento del lenguaje natural, y a un ritmo impensable en las décadas precedentes. Estos ha sido posibles gracias principalmente a los avances tecnológicos que han aumentado enormemente nuestra capacidad de almacenamiento y procesado de datos. La revolución informática ha dado lugar a un cambio radical de paradigma y metodología en la lingüística computacional, permitiendo sustituir los antiguos modelos basados en reglas gramaticales y formalismos teóricos con nuevos algoritmos estadísticos basados en el aprendizaje automático sobre grandes córpora de información lingüística.

La lista de nuevas aplicaciones de la tecnología lingüística no para de crecer. La necesidad de compartir, procesar y aprovechar el volumen creciente de información digital ha dado lugar a nuevos métodos basados en la lingüística de corpus en todos los ámbitos: la traducción automática y la comprensión del lenguaje natural para facilitar la comunicación, la respuesta automática a preguntas y el resumen de textos para agilizar la búsqueda de información, la detección de la opinión para fines comerciales y políticos, y muchos más.

Otro campo de aplicación de reciente aparición es el de la *minería de argumentos*. Motivado inicialmente por la necesidad de identificar y extraer argumentación de los documentos jurídicos, ya se está aprovechando para otros usos tales como el análisis de la opinión, la toma de decisiones, el análisis del discurso y el diagnóstico médico. Una variante de la minería de argumentos que todavía no ha sido explorada es la *minería de falacias* en la argumentación – es decir, la detección de argumentación con fallos de razonamiento, sea por invalidez lógica o por incoherencias en su contenido. Esto tiene mucho interés para la verificación de la información en aplicaciones como el análisis del discurso, el periodismo, el análisis político y sociológico, y otras en las que es necesario comprobar la validez de la argumentación para estudiar y evitar fenómenos como las "noticias falsas" y la manipulación de la opinión pública que son tan prevalentes hoy en día.

En el presente trabajo se propone una primera aproximación al problema de la minería de falacias. A partir del análisis de un corpus con texto argumentativo conteniendo múltiples ejemplos de argumentos falaces, se propondrán criterios para la correcta identificación y anotación de las falacias en el corpus y se definirán patrones lingüísticos que podrían servir para su detección mediante algoritmos computacionales basados en el aprendizaje. El trabajo se centrará en una clase concreta de falacia, la denominada "hombre de paja", que se presenta como un buen candidato para un primer abordaje del problema y además presenta múltiples instancias en el corpus seleccionado.

A continuación, se explicarán más en detalle estos conceptos y se resumirá el estado del arte en el campo de la minería de argumentos. Seguidamente, se concretarán los objetivos, el alcance del trabajo y la metodología propuesta. Finalmente, se detallará el análisis realizado y se presentarán los resultados obtenidos.

2 LA MINERÍA DE ARGUMENTOS

Un área de investigación relativamente nueva de la lingüística de corpus es la *minería de argumentos*. El objetivo principal de la minería de argumentos es diseñar métodos capaces de analizar la argumentación de las personas desde un punto de vista computacional. Se trata de identificar y extraer argumentos de córpora textuales con el fin de obtener datos estructurados para modelos computacionales de la argumentación y motores de razonamiento [Habernal y Gurevych 2016]. Aunque tiene relación con la minería de la opinión, donde se intenta saber *qué* piensan las personas sobre algo, con la minería de argumentos se aspira a saber *por qué motivos* piensan lo que piensan, desvelando sus *procesos de razonamiento* y no solo sus opiniones, juicios y sentimientos.

Una de las motivaciones iniciales de la minería de argumentos fue la de facilitar el análisis de documentos jurídicos. Su utilidad potencial en este campo es evidente dado que los procesos judiciales suelen requerir el análisis laborioso de grandes volúmenes de jurisprudencia para encontrar argumentación sobre casos pasados relevantes para casos actuales. Esta utilidad se extiende a otros campos donde se requiere la toma de decisiones. Los argumentos de otros en foros de discusión podrían ayudarnos a decidir si enviar a nuestros hijos a la escuela pública o privada, por ejemplo, o para posicionarnos sobre temas políticos, o para elegir un determinado modelo de coche o electrodoméstico. Disponer de herramientas capaces de buscar y resumir argumentos facilitaría la toma de decisiones sobre temas complejos o controvertidos y ayudaría a evitar la "sobrecarga de información" que habitualmente sufrimos al buscar información en línea. Otros campos de aplicación son el análisis del discurso, el *business intelligence*, el diagnóstico médico, la mediación y la educación.

2.1 Modelos de argumentación

El arte de la argumentación ha sido objeto de estudio desde los primeros trabajos de Aristóteles que remontan al siglo 4 A.C. Según su definición más básica, un *argumento* puede considerarse una *proposición* o *tesis* apoyada por *razones* o *premisas*. Dicho de otra forma, es un razonamiento mediante el cual se intenta probar, refutar o justificar una proposición o tesis. Desde un punto de vista pragmático, la *argumentación* es una variedad discursiva con la cual se pretende defender una opinión y persuadir de ella a un receptor mediante pruebas y razonamientos – es decir, argumentos. Es el arte de *persuadir* a otros para que piensen o actúen de una forma determinada mediante la presentación de argumentos. Hace un uso intensivo de la *retórica*, que es el arte de formular argumentos efectivos y persuasivos. Hoy en día, la argumentación es un campo de investigación multidisciplinario que estudia los procesos de debate y razonamiento abarcando y enlazando diversas áreas tales como la lógica, la filosofía, la lingüística, la retórica, el derecho, la psicología y la informática [Lippi y Torroni 2014].

Los trabajos de minería de argumentos suelen basarse en algún modelo teórico de la argumentación. Se han propuesto muchos modelos significativos, pero no existe ningún modelo unificado y universalmente aceptado de lo que constituye un argumento. La mayoría de los modelos de argumentación caen dentro de uno de dos enfoques principales – los modelos *abstractos* o los modelos *estructurados*. Los modelos abstractos consideran que un argumento es una entidad atómica sin estructura interna, mientras que los modelos estructurados proponen alguna estructura interna para cada argumento, descrito mediante algún formalismo de representación del conocimiento. Estos últimos son los que típicamente se emplean en la minería de argumentos, cuyo objetivo se plantea como la identificación y extracción de los diferentes elementos de los argumentos del lenguaje natural, por lo que éstos han de definirse con alguna estructura. Uno de estos modelos empleado con frecuencia en su forma original o modificada es el de Toulmin [1958], cuyo esquema se muestra en la Figura 1.

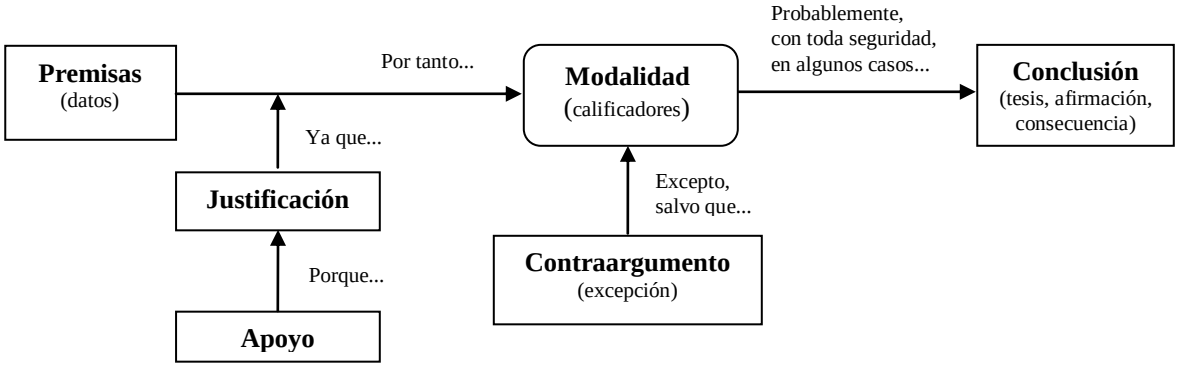


Figura 1: Modelo de la argumentación de Toulmin. Las casillas representan los diferentes componentes de un argumento, posiblemente marcados por conectores léxicos, y las flechas indican las relaciones implícitas entre componentes y el flujo lógico de la argumentación desde las premisas hasta la conclusión.

Un ejemplo de un argumento anotado según el modelo de Toulmin podría ser el siguiente [Habernal y Gurevych 2016]:

> [Harry nació en Bermuda.]Premisa **Ya que** [Un hombre nacido en Bermuda generalmente será un ciudadano británico.]Justificación **De acuerdo con** [Los siguientes estatutos y provisiones legales: (...)]Apoyo **Por lo tanto,** [presumiblemente]Modalidad **A menos que** [Sus dos padres fueran extranjeros]Contraargumento [Harry es un ciudadano británico.]Conclusión

El modelo de Toulmin pretende ser un modelo "ideal" que recoge todos los elementos básicos que podrían aparecer en cualquier tipo de argumento. En el discurso real, sin embargo, frecuentemente algunos de estos componentes no están presentes o se entienden de forma implícita. Como mínimo habrá una o más *premisas*, pero a menudo no habrá *justificación*, *apoyo*, *contraargumento* o *modalidad* explícitos y a veces incluso la *conclusión* puede ser implícita. Los diferentes componentes tampoco estarán marcados necesariamente por conectores léxicos explícitos sino por otras mecanismos léxicos, sintácticos o semánticos.

Es importante resaltar que el modelo de Toulmin se centra en los argumentos a nivel *monológico*, es decir, la estructura de un único argumento expresado por un único orador. Otros modelos intentan representar los argumentos a nivel *dialógico* o *retórico* en el contexto de un diálogo. La idoneidad de un modelo u otro para una aplicación de minería de argumentos dependerá del tipo y ámbito de discurso del texto estudiado y del enfoque propuesto. Para un estudio basado en un modelo monológico como el de Toulmin, el objetivo de la tarea de minería consiste en llegar a identificar computacionalmente los diferentes componentes de cada argumento de cada uno de los diferentes oradores, por separado, en un texto. Este es el enfoque más indicado para el presente trabajo puesto que, en el corpus estudiado, los argumentos falaces de interés suelen aparecer en las respuestas individuales de un único orador.

2.2 Métodos

Los trabajos de Habernal y Gurevych [2016] y Lippi y Torroni [2016] presentan una visión global y actualizada del estado de arte en el campo de la minería de argumentos, desde los campos de aplicación, los varios modelos empleados, y las tareas y subtareas abordadas hasta los córpora y técnicas de anotación, los métodos de aprendizaje ensayados y los resultados obtenidos. En términos generales, abordar un problema de minería de argumentos típicamente comprende la siguiente secuencia de tareas:

1. Elegir el modelo de argumento más adaptado al corpus y la aplicación.
2. Anotar los componentes argumentales en el corpus de acuerdo con el modelo elegido.
3. Definir y generar un conjunto de rasgos para entrenar un clasificador estadístico.
4. Elegir y entrenar un clasificador estadístico para que aprenda a asociar los rasgos de entrada con los patrones de salida (anotados en el corpus).

En las palabras de Lippi y Torroni [2016], "todos los enfoques de minería de argumentos propuestos hasta la fecha pueden describirse como sistemas de flujo multi-etapa, cuyos datos de entrada son un documento de texto en lenguaje natural y cuyos datos de salida son un documento anotado donde quedan anotados los argumentos y sus componentes. Cada etapa aborda una sub-tarea de todo el problema de minería de argumentos empleando uno o más metodologías y tecnologías de aprendizaje automático y procesado del lenguaje natural". Este proceso se ilustra en la Figura 2.

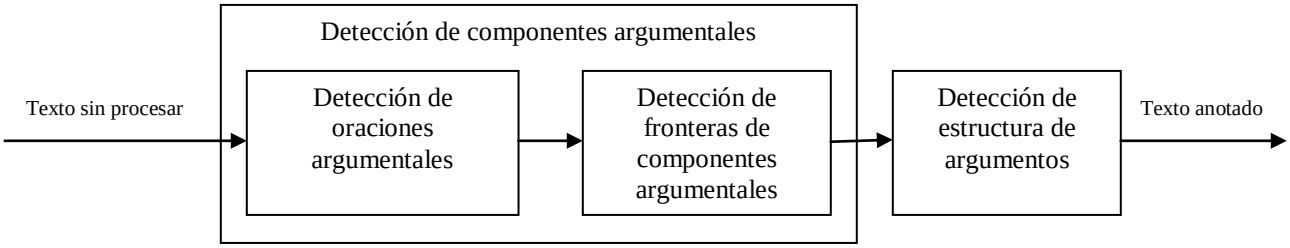


Figura 2: Diagrama de flujo de un sistema de minería de argumentos.

La detección de los componentes argumentales a menudo se divide en dos subtareas – la detección de todas las oraciones que contienen parte de un argumento, y la detección del inicio y final de los componentes argumentales en cada oración. La primera tarea es un problema de *clasificación* y en la mayoría de los trabajos se emplean clasificadores clásicos y conocidos tales como Máquinas de Soporte Vectorial (SVM), Regresión Logística, *Naïve Bayes*, Entropía Máxima, Redes Neuronales Recurrentes, Probabilidad Máxima, Árboles de Decisión, Bosques Aleatorios y otros. Los clasificadores se entrenan de forma supervisada mediante una colección de ejemplos etiquetados donde, para cada ejemplo, se proporciona una representación del texto a clasificar (un vector de rasgos) junto con la clase asociada (la etiqueta asignada en el proceso de anotación).

Según los mismos autores, se han comparado resultados con diferentes clasificadores pero no hay evidencias claras indicando cuáles son más preferibles. En cambio, casi todo el esfuerzo se ha dedicado a la creación de conjuntos de rasgos altamente especializados, por lo que se ha demostrado que el elemento

clave para obtener buenos resultados es la elección de rasgos y no el algoritmo de clasificación. Pero allí también, muchos trabajos típicamente emplean rasgos clásicos para la representación del texto. Entre los rasgos más empleados se encuentran los siguientes:

- *Bag of Words* (BoW) y sus variantes con bigramas y trigramas.
- Conocimiento proveniente de ontologías, tesauros y bases de datos léxicos como WordNet.
- Información gramatical proveniente de *parsers* sintácticos y *taggers* de POS.
- Información sobre la puntuación, tiempos verbales y marcadores discursivos.
- Rasgos generados con predictores externos tales como valoraciones de subjetividad, detectores de sentimiento o sistemas de reconocimiento de nombres propios.
- Rasgos extraídos manualmente del *RST Treebank* [Carlson 2002].
- Información específica del contexto como el tema, frases clave o conectores.

La segunda tarea – detección de las fronteras de los componentes argumentales – es un problema de *segmentación* o etiquetaje secuencial, en la que se debe asignar una clase (etiqueta) a cada palabra de una oración indicando su pertenencia o no a un componente argumental. A este problema se han aplicado métodos tales como Campos Aleatorios Condicionales, Modelos de Markov, y otros similares.

2.3 Detección de argumentos falaces

Todos los trabajos de minería de argumentos hasta la fecha se han centrado en el modelado y detección de argumentos tomados como lógicos y válidos – ha habido poca o casi ninguna investigación sobre el modelado y minería de argumentos falaces, tal como corroboran Habernal y Gurevych [2016] en su repaso reciente del estado del arte. En un intento de tener mínimamente en cuenta esta clase de argumentos en sus trabajos, estos mismos autores los etiquetaron en su totalidad simplemente como argumentos que "apelan a la emoción", pero sin intentar modelizar sus distintos componentes ni tampoco distinguir entre distintas clases de falacia. A falta de un modelo específico y debido al alto desacuerdo entre anotadores, decidieron excluir estos argumentos de su análisis.

Un intento reciente de explicar los argumentos falaces según el modelo de Toulmin se encuentra en el trabajo de Pineau [2013]. Su análisis se limita a identificar aquellos componentes argumentales del modelo que presentan debilidades o insuficiencias en algunas clases de falacia, sin especificar cómo éstos se podrían identificar.

En la próxima sección se explican el concepto de falacia y algunas clases de falacia. A continuación se comenta con más detalle la falacia del hombre de paja y se analizan algunas de sus características que podrían emplearse para su modelado y detección.

3 LA MINERÍA DE FALACIAS

3.1 Las falacias lógicas

Cuando el razonamiento de un argumento carece de validez lógica, se llama una *falacia*. Un error de razonamiento puede ocurrir tanto de forma accidental como deliberada. El uso consciente y habitual de las falacias como recurso retórico es frecuente en la argumentación cuando el enfoque se centra más en la persuasión que en el correcto razonamiento. En este contexto, se puede entender una falacia como un argumento que parece mejor de lo que es en realidad.

Es importante resaltar que la presencia de una falacia en un argumento no significa que no pueda ser persuasivo. Es más, muchas personas quedan persuadidas por argumentos falaces porque no llegan a identificar la falacia en el argumento. Las falacias a menudo son el último intento de un orador mal informado o poco preparado que se encuentra sin nada mejor que decir [University of Minnesota 2016]. Por el contrario, un orador astuto y experimentado puede emplear las falacias como un arma persuasiva para manipular un debate a su favor, aprovechándose de la ignorancia e ingenuidad de su audiencia. A menudo, los oyentes de un debate no están informados sobre los temas en discusión. De hecho, para tener éxito, un argumento falaz requiere que la audiencia sea ignorante o mal informada sobre el argumento original. Esto significa que incluso un orador ignorante y mediocre puede crear la falsa impresión de construir un argumento inteligente y coherente al emplear una falacia [Khare 2016]. Aunque el empleo hábil de las falacias puede acrecentar los méritos aparentes de un orador, su uso descuidado o excesivo puede delatar la invalidez de sus argumentos y contribuir a su descrédito.

El estudio de las falacias es tan antiguo como el estudio de la argumentación en sí. El mismo Aristóteles hizo una descripción y clasificación de las falacias en su obra *Refutaciones Sofísticas* [Hansen 2015]. Más recientemente, algunos autores han identificado y nombrado más de 125 tipos de falacia [University of Minnesota 2016]. Debido a su gran variabilidad de estructura y aplicación, no existe un acuerdo general sobre la manera de clasificarlas pero, a grandes rasgos, pueden dividirse en dos principales grupos

según si su invalidez radica en su estructura lógica (falacias formales) o en su contenido semántico (falacias informales). A diferencia de las falacias formales, las falacias informales no se prestan al análisis con la lógica formal y requieren un tratamiento basado en criterios semánticos. Éstas últimas son tal vez las más numerosas y pueden subdividirse en categorías tales como la relevancia lingüística por omisión, la relevancia por intrusión y la relevancia por presunción. Otras categorías comunes de falacias informales son las generalizaciones falsas y las pistas falsas (*red herrings*) que, a su vez, tienen muchos subtipos [Wikipedia 2018]. Algunas de las falacias informales más comunes y que aparecen con frecuencia en el contexto de la argumentación política incluyen las siguientes:

- **Pendiente resbaladiza** (*slippery slope*): También conocida como **efecto dominó**. Sugiere que una acción relativamente pequeña iniciará una cadena de eventos relacionados que culminarán en un evento posterior predecible e indeseado, por lo que se debe evitar la primera acción.
- **Conclusión irrelevante** (*missing the point*): Se presenta un argumento que puede ser por sí mismo válido, pero que no tiene que ver con la proposición que se debería probar o concluir.
- **Supresión de pruebas** (*cherry picking*): Consiste en seleccionar y señalar casos o datos que parecen confirmar la proposición deseada, a la vez que se ignora una porción importante de casos o datos que podrían contradecir esa posición.
- **Generalización apresurada** (*hasty generalization*): Ocurre al inferir una conclusión general en base a una muestra insuficiente de pruebas.
- **Pista falsa** (*red herring*): Consiste en distraer la atención del tema tratado introduciendo un argumento nuevo no relevante pero más fácil de tratar por el hablante.
- **Hombre de paja** (*straw man*): Consiste en caricaturizar los argumentos o la posición del oponente tergiversando, exagerando o cambiando el significado de sus palabras para justificar un ataque lingüístico o dialéctico.
- **Apelar a las emociones (miedo)** (*appeal to emotion*): Consiste en tratar de manipular las emociones del público en lugar de usar razonamientos válidos.
- **Ad hominem**: Consiste en atacar a la persona que emite un argumento, desacreditándole para que los demás no lo tengan en consideración.
- **Falacia de la causa simple** (*fallacy of the simple cause*): También conocida como **sobresimplificación causal**. Ocurre cuando se argumenta bajo la suposición de que existe una sola causa para un resultado cuando en realidad puede haber un conjunto de causas.

3.2 Las falacias en el debate político

Un ámbito comunicativo que se presta especialmente a la aparición de falacias argumentativas es el de los debates políticos. Un debate es una discusión formal, de carácter argumentativo, en el que dos o más personas exponen sus puntos de vista sobre un determinado tema polémico bajo la guía de un moderador. Cuando el objeto de discusión son temas políticos sensibles que pueden afectar las vidas tanto de los oradores como del público, los debates a menudo se cargan de emotividad y toman la forma de una batalla retórica donde los hablantes despliegan todos sus recursos para convencer a los oyentes de su posición. El debate político es omnipresente en nuestra sociedad actual y no solo entre candidatos políticos – participan también académicos, periodistas, analistas y el público general en foros diversos tales como mítines, congresos, tertulias televisadas, escenarios públicos y foros en internet.

En los debates políticos, el enfoque se centra en los aspectos *retóricos* y no *dialécticos* de la argumentación. Es decir, que el objetivo principal de los oradores es conducir el discurso y la opinión del público hacia sus propias ideas e intereses y no tanto presentar argumentos impecables. En este contexto, no es de extrañar que tiendan con frecuencia a descuidar su compromiso con el razonamiento crítico y opten por emplear estrategias retóricas injustas y engañosas a fin de manipular el discurso a su favor y persuadir la audiencia de sus ideas. El componente inmediato y espontáneo de un debate público, a menudo con fuertes limitaciones de tiempo, requiere un cierto grado de improvisación, lo cual contribuye en gran medida a esta tendencia. Desde un punto de vista psicológico, una falacia en la comunicación política puede definirse como un error de razonamiento empleado con fines engañosos. Las falacias informales son tan comunes en los debates políticos que han sido objeto de muchos estudios [Roitman 2017; Zurloni y Anolli 2016].

En los debates presidenciales norteamericanos del 2016 entre Clinton y Trump se pueden identificar ejemplos de falacias de varios tipos empleadas por ambos candidatos. Veamos a continuación algunos ejemplos (todos de [Chiasma 2016]):

(1) "Yes, I'm very embarrassed by it. I hate it. But it's locker room talk, and it's one of those things. I will knock the hell out of ISIS. We're going to defeat ISIS. ISIS happened a number of years ago in a vacuum that was left because of bad judgment. And I will tell you, I will take care of ISIS." (Donald Trump)

En (1) Trump utiliza una **pista falsa**, la lucha contra ISIS, para distraer la atención del argumento principal que trata de sus comentarios en un video publicado en 2005.

(2) *"So I believe that this election has become in part so—so conflict-oriented, so intense because there's a lot at stake. This is not an ordinary time, and this is not an ordinary election. We are going to be choosing a president who will set policy for not just four or eight years, but because of some of the important decisions we have to make here at home and around the world, from the Supreme Court to energy and so much else, and so there is a lot at stake. It's one of the most consequential elections that we've had."* (Hillary Clinton)

En el ejemplo (2) de **apelación a las emociones**, Clinton crea un sensación falsa de urgencia y miedo para apoyar su argumento de que estas elecciones son diferentes.

(3) *"It was locker room talk, as I told you. That was locker room talk. I'm not proud of it. I am a person who has great respect for people, for my family, for the people of this country. And certainly, I'm not proud of it. But that was something that happened. If you look at Bill Clinton, far worse. Mine are words, and his was action. His was what he's done to women. There's never been anybody in the history politics in this nation that's been so abusive to women. So you can say any way you want to say it, but Bill Clinton was abusive to women."* (Donald Trump)

El ejemplo (3) es una falacia **ad hominem**, donde se observa cómo Trump ataca a Bill Clinton, el marido de su oponente, en lugar de responder al argumento planteado contra él sobre su trato a las mujeres.

(4) *TRUMP: Oh, you didn't delete them?*
CLINTON: It was personal e-mails, not official.
TRUMP: Oh, 33,000? Yeah.
CLINTON: Not — well, we turned over 35,000, so…"

En (4) cuando Trump cuestiona a Clinton sobre los 33.000 emails que borró, responde diciendo que entregaron 35.000 emails, lo que constituye una **conclusión irrelevante** ya que, sea cierto o no, no responde a la pregunta original.

(5) *"We have to get rid of the lines around the state, artificial lines, where we stop insurance companies from coming in and competing, because they want — and President Obama and whoever was working on it — they want to leave those lines, because that gives the insurance companies essentially monopolies. We want competition. You will have the finest health care plan there is. […] You're going to have plans that are so good, because we're going to have so much competition in the insurance industry."* (Donald Trump)

En el ejemplo (5), Trump argumenta que al eliminar las restricciones fronterizas a las empresas de seguros, la competencia dará lugar al mejor sistema de salud posible, lo cual es una **sobresimplificación causal** de un problema mucho más complejo (niveles diversos de ingresos, patologías diversas, costes sanitarios, etc.).

(6) *"If you go with what Hillary is saying, in the ninth month, you can take the baby and rip the baby out of the womb of the mother just prior to the birth of the baby. Now, you can say that that's OK and Hillary can say that that's OK. But it's not OK with me, because based on what she's saying, and based on where she's going, and where she's been, you can take the baby and rip the baby out of the womb in the ninth month on the final day. And that's not acceptable."* (Donald Trump)

En (6) Trump le propina a Clinton un ataque **hombre de paja**, distorsionando sus argumentos y atribuyéndole una posición que no apoya en relación con el derecho al aborto, diciendo que ella estaría de acuerdo con arrancarle un bebé del útero de su madre al final de su embarazo. A continuación analizaremos más en detalle esta clase de falacia, que es la que se ha seleccionado como objeto de estudio del presente trabajo.

3.3 La falacia del hombre de paja

En un debate público entre candidatos presidenciales y otros altos cargos públicos, es habitual que la discusión tome la forma de una contienda en la que cada candidato intente por todos los medios convencer al público de los méritos relativos de sus ideas y propuestas, al mismo tiempo refutando y desacreditando las de su oponente. Esto se consigue retratando las propias ideas de la manera más favorable posible y las del oponente de la manera más desfavorable, frecuentemente por medio de la exageración, la caricaturización, la distorsión o cualquier otro mecanismo que sirva para polarizar las posiciones sobre los temas en cuestión. Por esto una de las falacias más comunes en este tipo de debates es el *hombre de paja*.

El esquema lógico de la falacia del hombre de paja es el siguiente [van Onseler 2012]:

1) El orador **A** afirma **P**
2) El orador **B** critica a **A** por afirmar **F** (distinto de **P**)
3) Por tanto, la afirmación de **A** es falsa.

La falacia del hombre de paja consiste en crear una representación distorsionada, exagerada, o ridiculizada de la posición y los argumentos de un oponente para facilitar su refutación. Toma su nombre por analogía con un muñeco de paja que, por su aspecto ridículo y su fragilidad, es mucho más fácil de derribar que un hombre real. Un ejemplo de este tipo de falacia sería cuando un político de la izquierda propone una medida como nacionalizar la industria eléctrica como un bien básico para los ciudadanos y otro político de la derecha le acusa de querer volver al comunismo y convertirnos en la Unión Soviética. Nacionalizar la electricidad no es volver a la URSS, sin embargo se puede combatir la idea de esa manera exagerada.

falacias.escepticos.es

Dado que el objetivo principal de un debate no suele ser necesariamente llegar a la verdad sobre los temas tratados sino simplemente ganar el debate, comentar la posición y los argumentos reales del oponente puede resultar demasiado complicado e incluso contraproducente. Una estrategia mucho más ventajosa consiste en presentar versiones distorsionadas, caricaturizadas o sobresimplificadas de sus argumentos, tal vez con un componente emocional persuasivo, cuya refutación sea fácil e incluso imperativa. Se trata de ganar adeptos, aunque haya que arrastrar los argumentos por el fango [Sanjul 2018]. Por ejemplo (7) (ejemplos 7 - 9 extraídos de [Corpus 2016]):

(7) *I don't want to rip families apart. I don't want to be sending parents away from children. I don't want to see the deportation force that Donald has talked about in action in our country.*
(Hillary Clinton)

Al atribuir opiniones irremediablemente inadecuadas o repugnantes al oponente, hace que las virtudes de los argumentos propios parezcan más obvias y distrae la atención de su debilidad y, a veces, incluso desvía el debate del tema central. Para funcionar y ser aceptado por la audiencia, primero, una falacia del

hombre de paja ha de estar relacionada al menos indirectamente con el tema en discusión. Segundo, ha de tener suficiente reclamo emocional para llamar la atención inmediatamente. Por ejemplo (8):

(8) This is a pattern, a pattern of divisiveness, of a very dark and in many ways dangerous vision of our country, where he incites violence, where he applauds people who are pushing and pulling and punching at his rallies. That is not who America is. (Hillary Clinton)

Por este motivo, a menudo se presta a *arquetipos* y *extremos* tales como los prejuicios, el racismo, la intolerancia y las generalizaciones. Su creador normalmente llegará a extremos para aparentar ser moralmente virtuoso y superior, denunciando al hombre de paja como malo, incorrecto o injustificable [van Onseler 2012]. Un buen ejemplo es el citado arriba sobre el tema del aborto, pero también queda ejemplificado con este otro (9) sobre el tema del control inmigratorio:

(9) But it is important for us as a policy, you know, not to say, as Donald has said, we're going to ban people based on a religion. (Hillary Clinton)

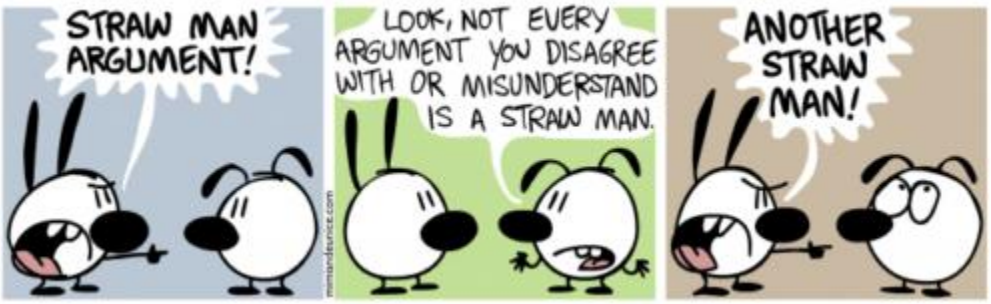
Realizado con habilidad, y sobre todo ante una audiencia ingenua o ignorante sobre los temas en cuestión, este tipo de argumento puede resultar convincente. Pero si el resultado es demasiado burdo u obvio, puede volvérsele en contra al que lo emplea y socavar su credibilidad. Por otro lado, el ataque de hombre de paja tiene el beneficio añadido de que deja al oponente en una posición desaventajada, puesto que se sentirá obligado a protestar y defenderse contra esta versión distorsionada de su postura, lo cual puede requerir más habilidad que argumentar en contra de una postura real.

3.4 El hombre de paja en el contexto de la negación

Otros estudios han mostrado como la *negación* se usa con frecuencia en los medios de comunicación y los debates políticos como una herramienta de argumentación. Por un lado, declara y distingue puntos de vista erróneos de aquellos que son válidos y polariza las tendencias de las dos posiciones políticas opuestas de un modo bastante económico: mientras se expresa el argumento propio se rechaza simultáneamente el punto de vista del oponente [Roitman 2017]. Además, a menudo resulta más cómodo para un candidato discutir en contra de los argumentos del otro que presentar su propio programa. Sin embargo, como cualquier herramienta, la negación como forma de refutación puede emplearse de un modo justo e informativo o, por el contrario, para desinformar y manipular el discurso. En un debate "ideal", la refutación se usa para poner en oposición dos o más opiniones y los participantes presentan

diferentes argumentos para apoyar una de las opiniones y persuadir a los espectadores de sus soluciones a los problemas. Pero en la realidad, y en función del tono del debate así como del estilo, la estrategia y las habilidades retóricas de los oradores, es habitual observarlos empleando métodos injustos de argumentación en la forma de falacias.

En relación con la negación, la falacia del hombre de paja es interesante porque se caracteriza por incluir una refutación en su propia *definición*. Un ataque de hombre de paja requiere no solo *construir* un argumento distorsionado sino también *derribarlo.* No es simplemente una falsa representación de la posición de otra persona, sino el empleo de esa falsa representación para refutar o criticar el argumento de esa persona en el contexto de un debate. En el argumento del hombre de paja, por definición, la posición distorsionada de un orador se usa para atacar, criticar o refutar el punto de vista de ese orador [Walton 1998]. Por lo tanto, ocurre potencialmente en contextos de negación oracional y otros tipos de refutación [Roitman 2017]. Este hecho puede servir como una pista lingüística para facilitar su detección y distinguirla de otras falacias similares tales como la *generalización apresurada* o la falacia *ad hominem* que también pueden construirse en base a generalizaciones, sobresimplificaciones y distorsiones sobre el oponente pero no requieren un ataque o refutación de sus argumentos.



La negación o refutación inherente en la falacia del hombre de paja también puede emplearse como pista para su detección computacional en procesos de minería, tal como se estudiará a continuación. Precisamente es de esperar que la presencia de pistas *lingüísticas* más o menos explícitas, previsibles y estructuradas en los textos simplifique en gran medida su detección computacional frente a otras falacias cuya identificación se tendría que basar mayormente en su evaluación *semántica* o con respecto a *información contextual* no presente en el texto. Por este motivo se ha considerado la falacia del hombre de paja un buen punto de partida para intentar una primera aproximación a la minería de falacias en este trabajo, sumado al hecho de que se encuentran abundantes ejemplos de este tipo de falacia en el corpus seleccionado.

3.5 Minería del hombre de paja

Parecería razonable abordar la minería de falacias como una variante de la minería de argumentos puesto que, por definición, las falacias son argumentos que carecen de validez, sea en su forma lógica (falacias formales) o en su contenido lingüístico o semántico (falacias informales). Se trataría, entonces, de identificar y caracterizar aquellos elementos de un argumento falaz que lo distinguen de un argumento válido y como un miembro de una clase determinada de falacias y adaptar un método de minería de argumentos para tener en cuenta esta información adicional. Estas adaptaciones quedarían reflejadas, en primer lugar, en los datos de entrenamiento extraídos del corpus, tanto en el etiquetaje de los argumentos a reconocer como patrones de salida como en el conjunto de rasgos seleccionados como patrones de entrada.

La lista de las diferentes clases de falacia es extensa, y las características que define y distingue cada clase de las demás son muy variadas y heterogéneas, por lo que no sería factible intentar definir criterios y métodos generales para detectarlas. Una estrategia más realista sería abordarlas por separado y, en caso de conseguir buenos resultados con una clase de falacias, intentar adaptar el método para tratar otras similares.

Para el caso tratado en el presente trabajo, conviene proponer un modelo teórico de la falacia del hombre de paja en el que se vean claramente cuáles son sus componentes argumentales y cuáles elementos lingüísticos podrían aparecer en las realizaciones de estos componentes argumentales en el discurso. Este modelo luego facilitaría tanto la identificación manual de las falacias en el corpus como la selección de patrones lingüísticos aptos para su minería computacional.

La falacia del hombre de paja generalmente se clasifica con las falacias de *relevancia*, un subgrupo de las falacias informales en las que se llega a una conclusión basada en argumentos que son irrelevantes al tema en cuestión. Concretamente, con un argumento hombre de paja se llega a una *conclusión válida* pero en base a *premisas irrelevantes*. En su análisis detallado de esta falacia, Walton [1996] presenta una serie de criterios y reglas normativas que permiten identificar correctamente un argumento hombre de paja y diferenciarlo de otras falacias relacionadas. Su estructura lógica puede representarse de la siguiente forma:

> Para dos oradores **M** y **N** y dos posiciones **P** y **F**:
> 1) **M** atribuye a **N** la posición **F**.
> 2) La posición de **N** no es **F** sino otra diferente, **P**,
> 3) **M** critica la posición **F** como si fuera la que realmente sostiene **N**.

Por lo tanto, la manera de detectar la falacia consistiría en, primero, observar que *M critica o refuta la posición de N* y, segundo, verificar que la posición criticada *F difiere de la posición P* apoyada en realidad por N.

El primer paso, la detección de la refutación, es similar a lo que se hace en la minería de argumentos "tradicional" – identificar una conclusión o afirmación que (opcionalmente) se apoya en una o más premisas, observaciones, evidencias u otros datos. Sin embargo, en este caso no se trata de una afirmación o conclusión positiva de la posición propia de M, sino de la *refutación* del orador M de la posición de N. La refutación podría tomar muchas formas, tales como una negación, una crítica, una descalificación, una valoración negativa, un contraste, o cualquier otro mecanismo que deje claro que M rechaza o está en desacuerdo con la supuesta posición F de N. Como mínimo, tendría que quedar evidente la *asociación* de F con el oponente N, y la *no asociación* de F con el hablante M.

El cumplimiento estricto del segundo requerimiento es más difícil ya que para poder comparar las dos posiciones, la supuesta F refutada por M y la real P sostenida por N, es necesario conocer P. La posición F argumentada por M ya será evidente puesto que es el objeto de su refutación, pero la posición real P de N sobre el tema en discusión no necesariamente forma parte del segmento de diálogo que contiene la falacia – puede haberse afirmado o citado mucho antes en el discurso o incluso no estar presente del todo. Idealmente, tomaríamos las palabras exactas y explícitas de N sobre el tema en cuestión durante el curso del diálogo como evidencias de sus compromisos al respecto, pero si no están presentes solo podemos detectar la falacia si esta información se entiende de forma implícita o si recurrimos a nuestra experiencia y nuestros conocimientos de fondo sobre lo que piensa N y las posiciones que representa.

Pero podemos superar esta problemática recordando que la falacia del hombre de paja no es una mera *representación falsa* o *distorsión* de la posición del oponente sustituyéndola por una versión diferente cualquiera, sino por una versión *exagerada*, *inverosímil*, *absurda*, *extrema*, *deplorable*, *caricaturizada*, *ridiculizada*, *sobregeneralizada*, o *sobresimplificada*. Por ejemplo, no sería un ataque hombre de paja si M dijera "N está en contra de la pena de muerte" cuando en realidad está a favor, pero sí lo sería si dijera "N quiere dejar libres a los asesinos en serie". El contenido semántico mismo de F llamará la atención por su inaceptabilidad en alguno de estos aspectos incluso sin conocer la posición real de N, por lo que *la comparación entre la falsa posición F y la posición real R es prescindible*. Por lo tanto, *el mismo contenido lingüístico de F debería contener todas las pistas necesarias para su identificación como parte de una falacia hombre de paja*.

Resumiendo, la existencia de una falacia del hombre de paja en un texto vendría indicada con una cierta probabilidad por la presencia simultanea de los siguientes elementos discursivos:

- La *asociación* de la proposición F con el oponente N.
- La *no asociación* de F con el hablante M.
- La *refutación* de la proposición F por parte del hablante M.
- El carácter *inaceptable* de la proposición F.

La realización de estos elementos, a su vez, implicaría la presencia probable de los siguientes elementos lingüísticos:

- Nombres y pronombres correspondientes a los interlocutores M y N.
- Lenguaje asociando N con una proposición F (afirmación, apoyo, valoración, voluntad, etc.).
- Lenguaje disociando M de F (negación, rechazo, refutación, descalificación, crítica, etc.).
- Lenguaje en F indicativo de exageración, inverosimilitud, absurdidad, extremismo, deplorabilidad, caricaturización, ridiculez, sobregeneralización, sobresimplificación, etc.

Por lo tanto, algunos rasgos lingüísticos que podrían servir para la detección de estos elementos indicadores de la falacia podrían ser:

elemento a detectar	**rasgos**
nombres y pronombres:	1-grama + POS
asociación N con F:	distancia entre N y F + n-gramas + POS
disociación M con F:	distancia entre M y F + n-gramas + POS
semántica de F:	n-gramas + polaridad + sintaxis + conocimientos de dominio

Estas listas de elementos discursivos y rasgos se proponen como guía para el proceso de anotación del corpus y la selección de rasgos. El análisis detallado del corpus y las falacias encontradas en ello permitirá verificar su utilidad e identificar otros posiblemente más adecuados.

4 OBJETIVOS Y ALCANCE

El objetivo del presente trabajo es hacer una primera aproximación a la minería de una clase de falacias en un corpus, concretamente la falacia del hombre de paja. El trabajo se centrará en abordar la primera y tal vez más crítica tarea en este proceso que es definir *criterios para la correcta identificación y anotación de la falacia en el corpus*, y proponer los *rasgos lingüísticos más adecuados que podrían servir de patrones de entrenamiento* para un algoritmo de detección automática. La selección y ensayo de dichos algoritmos queda fuera del alcance del trabajo y se propone como objeto de estudio futuro.

Como corpus del estudio se emplearán las transcripciones de los tres debates presidenciales norteamericanos del 2016 entre Donald Trump y Hillary Clinton, un texto argumentativo y relativamente formal en la que se ha verificado que hay un número significativo de esta clase de falacias. Mediante la revisión manual del texto se identificarán y se extraerán todos los ejemplos de la falacia del hombre de paja para su posterior análisis. Siguiendo la metodología que se describe a continuación, se definirán criterios de identificación y anotación de las falacias y se propondrán patrones lingüísticos de posible utilidad para su minería.

5 METODOLOGÍA

5.1 Corpus

Hubo tres debates presidenciales antes de las elecciones norteamericanas del 2016: el primero fue moderado por Lester Holt de la NBC y se celebró en la Universidad de Hofstra el 26 de septiembre, el segundo fue moderado por Anderson Cooper de la CNN y Martha Raddatz de ABC News y se celebró en la Universidad de Washington el 9 de octubre, y el tercero fue moderado por Chris Wallace de Fox News y se celebró en la Universidad de Las Vegas el 19 de octubre. Durante cada debate, los moderadores formulan preguntas a los dos candidatos presidenciales sobre los temas políticos en discusión, quienes responden por turnos de 2 minutos cada uno. Tanto la modalidad del discurso (debate político en forma de contienda) como su estructuración formal aseguran la producción de respuestas argumentativas por parte de los dos oradores en las que emplearán con toda probabilidad algunos recursos retóricos tales como las falacias. La transcripción de estas respuestas (excluidas las preguntas) constituyen el corpus del presente trabajo.

Para este estudio se empleó una versión del corpus [Corpus 2016] que ya fue compilado y descrito para un trabajo anterior del mismo grupo de investigación del CLiC (Centre de Llenguatge i Computació) de la Universitat de Barcelona [Badertscher 2017]. Algunas de sus estadísticas más destacables se recogen en la Tabla 1.

	Clinton	Trump
Total respuestas	226	342
Total palabras (tokens)	19523	22416
Total types	2528	2139
Types/tokens	0.129	0.095
Total oraciones	1200	2025
Media palabras / respuesta	86.4	65.5
Media palabras / oración	16.3	11.1
Mediana palabras / oración	13	8
Media letras / palabra	4.4	4.3
Mediana letras / palabra	4	4

Tabla 1: Estadísticas destacables del corpus para los dos oradores (Hillary Clinton y Donald Trump).

Cabe destacar que aunque Trump pronuncia un número mayor de respuestas y palabras totales que Clinton, Clinton emplea un número significativamente mayor de palabras diferentes (*types*), de palabras por oración y de palabras por respuesta que Trump, lo cual sugiere que Clinton habla con un vocabulario más variado y con mayor complejidad sintáctica que Trump. Esto previsiblemente irá acompañado de un mayor dominio por parte de Clinton de algunos recursos retóricos como las falacias, por ejemplo, como se verá en el análisis a continuación.

5.2 Preprocesado del corpus

Inicialmente, todas las respuestas de los dos candidatos se encontraban almacenadas en ficheros de texto separados, cuyos nombres son índices formados por las siglas del candidato, el número del debate, y la secuencia de la respuesta dentro de ese debate (ej.: HRC1.43). Para facilitar la revisión manual del corpus en el presente trabajo, mediante un *script* de UNIX se juntaron todas las respuestas de cada candidato en un único fichero para ese candidato, ordenadas secuencialmente y precedidas por el índice correspondiente. Estos ficheros fueron importados en documentos *MS-Word* para permitir su anotación mediante el formateo con colores y otros elementos.

5.3 Identificación manual del hombre de paja

La identificación de ejemplos de falacia en el texto se realizó de forma manual. La decisión de lo que constituye o no una falacia hombre de paja por lo general tenía un componente subjetivo importante y en algunos casos había múltiples interpretaciones posibles, sobre todo a la hora de juzgar el grado de distorsión, exageración, inverosimilitud, deplorabilidad, etc. de la falsa posición atribuida al oponente. El patrón básico buscado para señalar un candidato a falacia fue aquel descrito en la Sección 3.5, que consiste en identificar lo siguiente en la respuesta de un hablante[1]:

- Una *proposición falaz de contenido exagerado, inverosímil, absurdo, extremo, deplorable, caricaturizado, ridículo, sobregeneralizado o sobresimplificado*.

- Una clara *asociación con el oponente* del hablante de dicha proposición falaz, o la percibida *intención* de asociársela.

- Un claro *rechazo* de dicha proposición por parte del hablante que la *disocia* de sí mismo, mediante la negación, refutación, descalificación, crítica, etc.

Estando atento a este patrón, se pudieron identificar un número significativo de ejemplos para los dos interlocutores, aunque con un grado muy variable de certeza según el caso. Los criterios y supuestos empleados para guiar la identificación de falacias y facilitar su análisis posterior fueron los siguientes:

- *Se considera que una falacia está contenida en una única respuesta de un único interlocutor* o, como mucho, *en dos respuestas consecutivas del mismo interlocutor* si está claro que la segunda es continuación de la primera. Por ejemplo (10):

(10) So we have a very robust set of plans. And people have looked at both of our plans, have concluded that mine would create 10 million jobs and yours would lose us 3.5 million jobs, and explode the debt which would have a recession. [HRC1.28]

That can't - that can't be left to stand. [HRC1.29]

[1] De aquí en adelante se emplea la siguiente terminología: el "hablante" es el interlocutor que pronuncia la "proposición falaz" contra su "oponente" (**M**, **F** y **N** en la Sección 3.5, respectivamente).

- *No se contemplan casos en que una falacia abarque simultáneamente afirmaciones de ambos hablantes*, ni en sus intervenciones consecutivas al responder un hablante al otro, por ejemplo, ni en las no consecutivas en referencia a algo dicho por el oponente en un punto anterior del discurso.

- *Se presupone que cualquier afirmación inverosímil, deplorable, etc. asociada a un oponente es una distorsión de su posición real*, incluso pudiendo darse el caso contrario. Es decir, que si una afirmación parece inverosímil, extremo, deplorable, etc. se presupone que no representa fielmente lo que realmente piensa o apoya el oponente y por tanto forma parte de una falacia. Se descartan, sin embargo, aquellos casos en que la afirmación representa posiciones, acciones o hechos muy conocidos del oponente y por lo tanto no constituyen una distorsión. Por ejemplo, en el siguiente caso (11), las dos primeras frases del texto (en negro) hacen referencia a algo que realmente hizo el oponente y por tanto no forman parte de la falacia, mientras que la tercera frase (en verde) sí, ya que muestra signos claros de caricaturización y sobresimplificación:

(11) Donald has bought Chinese steel and aluminum. In fact, the Trump Hotel right here in Las Vegas was made with Chinese steel. So he goes around with crocodile tears about how terrible it is, but he has given jobs to Chinese steelworkers, not American steelworkers. [HRC3.37]

- *La asociación de la proposición falaz y rechazada con el oponente no ha de ser siempre explícita, sino que puede entenderse de forma implícita*, bien por sentido común, por conocimiento previo de la posición real del oponente y de sus afirmaciones, por la intención percibida del hablante de atribuir la proposición a su oponente, o por la estructura sintáctica o retórica del texto (por contraposición, contraste, etc.). En muchos casos, el simple hecho de que el hablante rechace una idea deja claro que no se trata de su propia posición sino la de su oponente. En el ejemplo (12), se entiende que las últimas dos frases (en verde) se asocian implícitamente a las políticas del oponente con respecto al papel de EE.UU. en la OTAN:

(12) And as far as Japan is concerned, I want to help all of our allies, but we are losing billions and billions of dollars. We cannot be the policemen of the world. We cannot protect countries all over the world. [DT1.108.2]

- *Los nombres y pronombres empleados para referirse al hablante y al oponente no siempre son los propios de esas personas* (ej. `I, he, Donald, Hillary`), sino que pueden ser referencias indirectas pero claras (`we, they, you`, etc.). En el ejemplo (13), se entiende que "`they`" incluye a Hillary Clinton puesto que formaba parte de la administración detrás de la política referida:

(13) And when they made that horrible deal with Iran, they should have included the fact that they do something with respect to North Korea. And they should have done something with respect to Yemen and all these other places. [DT1.104]

- *El rechazo de la proposición distorsionada no siempre se realiza mediante léxico explícito, sino que puede representarse de forma implícita* mediante el contraste o la yuxtaposición de proposiciones contrastantes. Por ejemplo (14):

(14) Donald has consistently insulted Muslims abroad, Muslims at home, when we need to be cooperating with Muslim nations and with the American Muslim community. [HRC1.68]

- *El léxico que más claramente asocia el hablante o el oponente con una proposición por lo general son verbos* (`thinks, says, has, wants to`, etc.) pero en algunos casos puede incluir léxico de otras categorías. En el siguiente ejemplo (15), la asociación se entiende por "`plan`" y no por el verbo "`has been`":

(15) Donald's plan has been analyzed to conclude it might lose 3.5 million jobs. [HRC3.32]

- *Las proposiciones que representan ideas y posiciones propias del hablante a menudo se anteponen o se posponen a la proposición falaz a modo de contraste y contraposición*, por lo que se consideran como un componente de la falacia y se les asigna una anotación específica. Por ejemplo (texto celeste en (16)):

(16) First, we have to build an economy that works for everyone, not just those at the top. [HRC1.2]

5.4 Anotación de proposiciones y elementos léxicos clave

Con estos criterios definidos, se procedió a identificar y anotar aquellos segmentos del corpus que parecían contener una falacia del hombre de paja. Para los objetivos planteados en este estudio, la anotación se realizó para cumplir con dos propósitos – primero, señalar y diferenciar las diferentes *proposiciones* que constituyen la falacia y, segundo, identificar *elementos léxicos y sintácticos* característicos de estas proposiciones que podrían servir de patrones lingüísticos para su detección automática.

La detección automática de una falacia puede plantearse como un proceso de reconocimiento de patrones a dos niveles. Si se entiende una instancia de falacia como un conjunto de unas clases determinadas de proposiciones, cada una de éstas identificable mediante una serie de patrones específicos, las cuales típicamente se presentan en un orden sintáctico determinado, entonces el proceso de reconocimiento consistiría en identificar en un primer nivel las diferentes proposiciones típicas de la falacia y, en un segundo nivel, detectar que se encuentran en un orden sintáctico característico de esta clase de falacia. Nuestra tarea de anotación, por lo tanto, consiste en señalar tanto los patrones que deseamos reconocer (el conjunto de proposiciones que componen la falacia) como los patrones léxicos y sintácticos que podrían servir para entrenar un algoritmo de reconocimiento automático.

El proceso de anotación se realizó por un solo anotador (el autor) únicamente con fines de realizar una primera aproximación al problema, y no constituye una anotación consensuada y definitiva para el entrenamiento de un sistema de reconocimiento automático. Por otra parte, la anotación se realizó solamente sobre aquellos segmentos del corpus conteniendo ejemplos de falacias del hombre de paja – con el objetivo de descubrir patrones de reconocimiento – y no sobre el resto del corpus.

A diferencia de la anotación para la *minería de argumentos*, donde se suele etiquetar la función de los diferentes componentes argumentales (premisa, justificación, apoyo, conclusión, etc.) de acuerdo con un modelo de argumentación, para la identificación de la falacia del hombre de paja se ha considerado suficiente anotar únicamente dos tipos de proposición – las *proposiciones falaces asociadas al oponente y rechazadas por el hablante*, y las *proposiciones asociadas a la posición del hablante* – sin necesidad de distinguir su función argumental. Para esto se utilizó las siguientes etiquetas[2]:

[2] En los ejemplos de falacias anotadas que aparecen en esta y las siguientes secciones, el texto anotado se muestra formateado con colores en lugar de etiquetas para simplificar su lectura y visualización, usando los colores correspondientes a cada etiqueta.

<falacia>	proposición falaz asociada con oponente
<posición>	proposiciones asociadas a la posición del hablante

La identificación de las proposiciones falaces (etiquetadas como <falacia>) requiere la detección de tres pistas importantes: el carácter falaz de su contenido semántico, elementos léxicos que expresan el rechazo de la proposición por parte del hablante, y la asociación con el oponente. El contenido semántico de las proposiciones falaces en la mayoría de los casos es demasiado complejo para ser detectado o representado con elementos léxicos aislados y requiere un análisis más elaborado, como se comentará más adelante. El rechazo o refutación de las proposiciones falaces, sin embargo, suele ser fácilmente identificable mediante elementos léxicos concretos, y se etiquetaron de la siguiente manera en el corpus:

<refutación>	léxico que rechaza/refuta la proposición falaz

Finalmente, la asociación explícita de una proposición con el hablante o el oponente, o su rechazo por parte del hablante, implica la presencia de referencias a estas personas y léxico que las asocia al contenido de la proposición. Por lo general (pero no en todos los casos), esto se realiza con el *nombre* o *pronombre* de la persona en el papel de sujeto de un *verbo* que expresa lo que esa persona piensa, siente, desea, dice, ha hecho, etc. con respecto al contenido de la proposición. Estos elementos se anotaron de la siguiente manera:

<hablante>	nombre/pronombre que hace referencia al hablante
<oponente>	nombre/pronombre que hace referencia al oponente
<asociación>	léxico asociando hablante/oponente con una proposición

Una vez identificadas y anotadas todas las falacias del corpus junto con sus elementos léxicos clave, se extrajeron del resto del texto del corpus y se juntaron en un único fichero para cada hablante junto con sus índices, para facilitar su estudio y manipulación de forma aislada.

5.5 Identificación de patrones sintácticos

El paso siguiente consistió en clasificar y agrupar las falacias de acuerdo con su estructura sintáctica – es decir, la secuencia en que aparecen las diferentes proposiciones y elementos léxicos clave que las componen[3] – para descubrir posibles patrones de secuencia (como parte de recursos discursivos, por

[3] Aquí no se trata de una estructura "sintáctica" propiamente dicha (que requería una representación jerárquica) sino de una estructura "secuencial". Esta representación aproximada de la sintaxis se ha creado aquí únicamente para facilitar la agrupación y ordenación de las falacias por las similitudes entre sus estructuras.

ejemplo). Para facilitar este proceso, se definió un código abreviado para representar las clases de proposiciones, los elementos léxicos clave que identifican las proposiciones, y otros elementos sintácticos, que fue el siguiente:

1	nombre/pronombre que hace referencia al hablante
3	nombre/pronombre que hace referencia al oponente
T	pronombre impersonal o demostrativo que hace referencia a F
F	proposición falaz asociada con oponente
R	rechazo/refutación de la proposición falaz
P	proposiciones asociadas a la posición del hablante
C	conector de contraste (*when*, *in contrast*, etc.)
.	punto entre dos frases
+	coordinador entre dos sintagmas (*and*, *but*, etc.)

Para cada falacia se creó una representación abreviada de su estructura sintáctica mediante la concatenación de estos códigos. Por ejemplo:

1**R**3**F**
1P**RF**
3F.3F+3**F**.1**R**.1P
1P+3**F**.T**R**
(etc.)

Teniendo estas representaciones abreviadas de todas las falacias identificadas, se procedió a agruparlas por sus similitudes estructurales y luego ordenarlas por orden de complejidad creciente. Los patrones sintácticos ordenados junto con el léxico clave identificado en todas las falacias se recogieron en las Tablas 2 y 3 del Anexo 1 para cada uno de los dos candidatos. En la próxima sección se analizarán los resultados obtenidos.

6 DISCUSIÓN Y RESULTADOS

En las Tablas 2 y 3 del Anexo 1 se recogen los datos recogidos para todos los ejemplos de falacia del hombre de paja identificados en el corpus para los dos candidatos, que fueron 57 en el caso de Hillary Clinton y 28 para Donald Trump. Su análisis revela algunos patrones claros candidatos a facilitar la identificación de los diferentes componentes de la falacia del hombre de paja. A continuación se

comentará cada uno de ellos y cómo podrían aprovecharse para el entrenamiento de un algoritmo de reconocimiento automático.

6.1 Identificación de la proposición falaz

Como ya se comentó en la Sección 5.3, la identificación de las proposiciones falaces requiere detectar tres pistas: *carácter inaceptable de su contenido*, su *asociación con el oponente* y su *rechazo por parte del hablante*. Mientras que la asociación y el rechazo se pueden detectar mediante léxico clave de categorías cerradas, como se detallará más abajo, la detección del carácter inaceptable del contenido semántico es más complicado puesto que requiere un análisis semántico de la proposición entera para determinar su grado de inverosimilitud, exageración, absurdidad, ridiculez, sobresimplificación, etc. Las posibles combinaciones de léxico que pueden expresar ideas de estas características son ilimitadas y en muchos casos requieren una evaluación semántica más compleja que simplemente la presencia de algunas palabras, bigramas o trigramas clave, aunque en otros casos podría ser suficiente. Por ejemplo, en (17) el tri-grama "`rip families apart`" de por sí ya tiene connotaciones de exageración e inaceptabilidad:

(17) I don't want to rip families apart. [HRC3.8]

En otros casos como (18) la falacia contiene varios n-gramas que por sí solos expresan ideas inaceptables, como "`hateful`", "`devisive campaign`", "`inciting of violence`", y "`brutal`":

(18) It's with him and with the hateful and divisive campaign that he has run, and the inciting of violence at his rallies, and the very brutal kinds of comments about not just women, but all Americans, all kinds of Americans. [HRC2.43]

Para estos casos, incluir información sobre la *polaridad de los n-gramas* en los rasgos de entrenamiento podría ser una estrategia útil para mejorar la detección de la falacia por un algoritmo de clasificación.

Sin embargo, hay muchos otros casos como (19), donde ninguno de los n-gramas "`more advantages`", "`advantages for people`" y "`very top`" de forma aislada dan una indicación de la inaceptabilidad del sintagma completo:

(19) Broad-based, inclusive growth is what we need in America, not more advantages for people at the very top. [HRC1.42.2]

En este caso sería necesario un *análisis sintáctico* para establecer la relación preposicional "for" entre "more advantages" y "people at the very top", junto con *conocimiento del dominio socio-político* para juzgar la inaceptabilidad de esta relación.

6.2 Identificación de la refutación

En cuanto al rechazo de la proposición falaz, en todos los casos salvo 3 para HRC y 2 para DT se identificaron elementos léxicos claros cuya función era la refutación explícita de la proposición falaz. La categoría del léxico depende de la construcción semántica del texto, pero en la mayoría de los casos son o incluye palabras de negación (not, no, never, nobody, etc.).

En muchos ejemplos como (20) y (21) con la construcción 1P**R**F (posición hablante + refutación + proposición falaz), la refutación se realiza contrastando una proposición de posición del hablante con la proposición falaz mediante un adverbio de negación (not, never, etc.), separando el verbo de la refutación:

(20) I also want to see more companies do profit-sharing. If you help create the profits, you should be able to share in them, not just the executives at the top. [HRC1.2.2]

(21) So I've tried to be very specific about what we can and should do, and I am determined that we're going to get the economy really moving again, building on the progress we've made over the last eight years, but never going back to what got us in trouble in the first place. [HRC1.7]

El léxico clave encontrado en estos casos incluía:

not, no, nobody, never, don't (imperative)

En otros casos con construcciones como 1**R**F (pronombre hablante + refutación + proposición falaz) o T**R** (pronombre impersonal + refutación), la refutación incluye una negación, un sujeto nominal o pronominal y un verbo, como en los ejemplo (22) y (23):

(22) I don't want to rip families apart. [HRC3.8]

(23) So let's fix what's broken about it, but let's not throw it away and give it all back to the insurance companies and the drug companies. That's not going to work. [HRC2.20]

Algunas de las pistas léxicas encontradas para estos casos incluyen las siguientes:

don't think, are not going to, don't want, will not, cannot, has no business, don't let, fought against, don't buy, let's not, am so disappointed, am not, should be, should be, haven't

En (24), (25) y (26) vemos construcciones conteniendo la secuencia **TR** (pronombre impersonal + refutación) donde la refutación se realiza con un pronombre impersonal (this, that, it, etc.) referido a la proposición falaz + verbo (generalmente "to be") + adverbio de negación y/o adjetivo negativo:

(24) People like Donald, who paid zero in taxes, zero for our vets, zero for our military, zero for health and education, that is wrong. [HRC2.27.2]

(25) I understand the border. She doesn't. She wants amnesty for everybody. Come right in. Come right over. It's a horrible thing she's doing. [DT2.48.2]

El léxico clave encontrado para estos casos incluye:

is no, is not, is not good, is wrong, can't be, shouldn't, has never been, should never be, is disgraceful, would be a terrible mistake, is a big problem, turned out to be wrong, is not OK, is not acceptable, is a horrible thing

En los dos casos (26) y (27) de HRC, la refutación se realiza empleando conectores de contraste (when, by contrast):

(26) Donald has consistently insulted Muslims abroad, Muslims at home, when we need to be cooperating with Muslim nations and with the American Muslim community. [HRC1.68]

(27) That is a plan that has been analyzed by independent experts which said that it could produce 10 million new jobs. By contrast, Donald's plan has been analyzed to conclude it might lose 3.5 million jobs... [HRC3.32]

En los 5 casos donde no hay léxico de refutación explícita, la refutación se entiende implícitamente por la yuxtaposición de la posición propia del hablante con la posición falaz atribuida del oponente, como en (28):

(28) I have costed out what I'm going to do. He will, through his massive tax cuts, add $20 trillion to the debt. [HRC3.34.2]

6.3 Asociación de proposiciones a interlocutores

Previsiblemente, el léxico encontrado en la mayoría de los casos que establece la asociación clara entre una proposición y un interlocutor consistía en el *nombre* o *pronombre* del interlocutor usado como sujeto o agente de un *verbo* que expresaba la implicación de ese interlocutor con el contenido de la proposición, que generalmente eran verbos que expresan *afirmación*, *pensamiento*, *creencia*, *sentimiento*, *ofrecimiento*, *aceptación*, *apoyo*, *valoración*, *voluntad*, *deseo*, *intención*, *compromiso*, *acción*, *realización*, etc. Los pronombres usados para expresar la posición propia del hablante incluyeron principalmente los de primera persona, tanto el singular como el plural:

I, my, mine, we, our

pero en dos casos – (29) y (30) – la relación con el hablante se entiende implícitamente por referencia a algo mencionado justo antes en el discurso usando un pronombre impersonal (`that`) o de tercera persona (`they`):

(29) That is a plan that has been analyzed by independent experts which said that it could produce 10 million new jobs. [HRC1.68]

(30) They need to have close working cooperation with law enforcement in these communities, not be alienated and pushed away as some of Donald's rhetoric, unfortunately, has led to. [HRC1.68.2]

De manera similar, los pronombres más comunes usados para referirse al oponente fueron referencias directas de tercera persona:

he, his, him, Donald, Donald Trump, she, her, Hillary, Hillary Clinton, Secretary Clinton, they

mientras que en unos pocos casos el discurso fue dirigido directamente al oponente empleando pronombres de segunda persona:

you, yours

En tres casos, Trump empleó el pronombre de primera persona plural "we" para referirse a un hecho indeseable del estatus quo, atribuible al oponente u otro oponente no identificado, como en (31) y (32):

(31) We're also letting drugs pour through our southern border at a record clip. At a record clip. And it shouldn't be allowed to happen. [DT2.48]

(32) I will bring our energy companies back. They'll be able to compete. They'll make money. They'll pay off our national debt. They'll pay off our tremendous budget deficits, which are tremendous. But we are putting our energy companies out of business. We have to bring back our workers. [DT2.88.3]

y en otros casos la referencia al oponente se hace indirectamente a través de otras personas (Barack Obama, Bill Clinton, Secretary Kerry) asociadas con el oponente o sus políticas, como en (33) y (34):

(33) If you look at Bill Clinton, far worse. Mine are words, and his was action. His was what he's done to women. There's never been anybody in the history politics in this nation that's been so abusive to women. So you can say any way you want to say it, but Bill Clinton was abusive to women. [DT2.13]

(34) And when asked to Secretary Kerry, why didn't you do that? Why didn't you add other things into the deal? One of the great giveaways of all time, of all time, including $400 million in cash. Nobody's ever seen that before. That turned out to be wrong. [DT1.104.2]

A diferencia de las posiciones propias del hablante que siempre incluyen un nombre o pronombre, la asociación de las posiciones falaces con el oponente no siempre incluía una referencia explícita al oponente. De hecho, en más de la mitad de los ejemplos de Clinton la refutación no hacía referencia directa al oponente, como en (35):

(35) Broad-based, inclusive growth is what we need in America, not more advantages for people at the very top. [HRC1.42.2]

En estos casos la asociación con el oponente se entiende de forma implícita del contexto discursivo. En cambio, en todas las refutaciones de Trump salvo tres está presente la referencia explícita al oponente por nombre o pronombre.

Los verbos encontrados en los ejemplos que asocian los interlocutores con las acciones o políticas mencionadas incluyen los siguientes:

> want, need, believes, hope, have to, will have, is doing, does, did, didn't do, has done, is not going to, have got to, caused by, made, said, is saying, tells, is telling, will tell, think, won't allow, agrees, got caught in, is blaming, is not OK with, is buying, letting, has talked about, brings up a point, will stand up for, is not that enthusiastic about, ran, feel, support, have set forth, has put forth, would consider, will be honest, has suggested, have tried, am determined, doesn't agree, inciting, exploit, admit, condemn, rejects, encouraged, refused, am proposing, let, know, didn't pay, will have accepted, may not have liked, are whining, has never apologized for, am proud of, didn't add, understand, can't tell, attacked, represented, has been seen laughing, will bring, is under siege by, is important to, can't work out, can't get together, can't sit them around

En tres casos la asociación se hace mediante el pronombre posesivo ('s, my) y un sustantivo + "*be*" como "argument is", "plan is" o "rhetoric has", como en (15) y (36):

(36) Well, within hours I said that I was sorry about the way I talked about that, because my argument is not with his supporters. It's with him and with the hateful and divisive campaign that he has run [HRC2.43]

6.4 Estructuras sintácticas

La clasificación y ordenación de las falacias por su secuencia composicional reveló algunos patrones secuenciales recurrentes, aunque con grandes diferencias entre las producciones de dos candidatos. En el caso de Clinton, había algunos patrones composicionales que se repetían con regularidad debido a su tendencia a presentar sus argumentos empleando recursos retóricos comunes. Algunos de los más empleados fueron (ejemplos (37) – (40)):

1**RF** refutación + falacia (12 casos)

(37) I don't want to be sending parents away from children. [HRC3.8.2]

1P**RF** posición + refutación + falacia (17 casos)

(38) We also, though, need to have a tax system that rewards work and not just financial transactions. [HRC1.3]

1P**R**3**F** posición + refutación + oponente + falacia (6 casos)

(39) I want to get everybody out of the shadows, get the economy working, and not let employers like Donald exploit undocumented workers, which hurts them, but also hurts American workers. [HRC3.10]

F.T**R** falacia + "." + pronombre impersonal + refutación (7 casos)

(40) The other day, I saw Donald saying that there were some Iranian sailors on a ship in the waters off of Iran, and they were taunting American sailors who were on a nearby ship. He said, you know, if they taunted our sailors, I'd blow them out of the water and start another war. That's not good judgment. [HRC1.71]

Sin embargo, hay una proporción importante de los ejemplos de Clinton que presentan patrones muy diversos, tales como 1P.1P.T**R**, 1P.1P.3**F**.**R**, 3**F**C1P, 3**F**.1P, etc.

Trump, en cambio, parece tender mucho menos al uso de recursos retóricos y en general habla con frases cortas, a veces incompletas, desestructuradas y sintácticamente simples, como ya observó

Badertscher [2017]. A menudo construye sus argumentos encadenando varias frases cortas y simples dando lugar a patrones estructurales repetidas, como en (41) y (42):

3F.3F.1P.1P.1P.**3F**.1P.**1R**F

(41) I was up in New Hampshire the other day. The biggest complaint they have - it's with all of the problems going on in the world, many of the problems caused by Hillary Clinton and by Barack Obama. All of the problems - the single biggest problem is heroin that pours across our southern border. It's just pouring and destroying their youth. It's poisoning the blood of their youth and plenty of other people. We have to have strong borders. We have to keep the drugs out of our country. We are - right now, we're getting the drugs, they're getting the cash. We need strong borders. We need absolute - we cannot give amnesty. [DT3.10.2]

1R.3F.3F.3F.3**F**.1**R**

(42) Well, I think I should respond, because - so ridiculous. Look, now she's blaming - she got caught in a total lie. Her papers went out to all her friends at the banks, Goldman Sachs and everybody else, and she said things - WikiLeaks that just came out. And she lied. Now she's blaming the lie on the late, great Abraham Lincoln. That's one that I haven't... [DT2.49]

A pesar de lo que parece una excesiva variedad y escasa previsibilidad en los patrones observados, hay algunos patrones generales que se repiten en casi todos los ejemplos y que podrían servir de pistas útiles para la detección de una falacia en lo que se refiere a su composición secuencial:

1) Siempre hay una proposición falaz **F** (por definición), con o sin referencia al oponente **3**.

2) Casi siempre hay una refutación explícita **R** en la misma proposición falaz o en la frase inmediatamente precedente o posterior. En los casos sin refutación explícita, hay una posición del hablante **1P** que hace la función de refutación por contraste.

3) Es frecuente encontrar una posición del hablante **1P** (o varias encadenadas) justo delante de la refutación o en la frase inmediatamente precedente o posterior, con función de contraste con la proposición falaz.

6.5 Modelado del hombre de paja

A partir del análisis realizado, ya es posible proponer un modelo aproximado de la estructura de la falacia del hombre de paja tal como se presenta en los ejemplos reales identificados en este corpus. El modelo propuesto se muestra en la Figura 3, donde los componentes argumentales típicos de la minería de argumentos (premisas, justificación, conclusión, etc.) se han sustituido por componentes de "posición", que representan afirmaciones (proposiciones) sobre lo que piensa o apoya el hablante o su oponente en el discurso, y el componente de "refutación", que representa el rechazo verbal de la posición falaz que el hablante atribuye a su oponente al pronunciar la falacia.

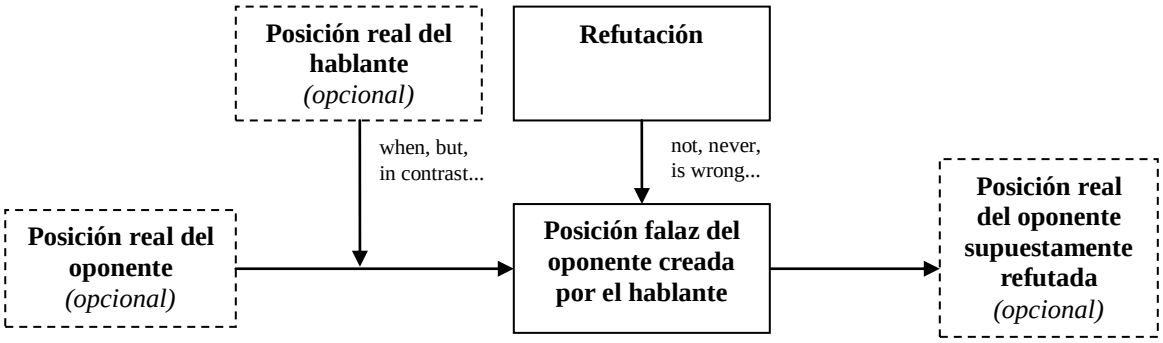


Figura 3: Modelo propuesto de la falacia del hombre de paja. Los componentes de "posición" representan afirmaciones/proposiciones en el discurso que expresan lo que piensan o apoyan los interlocutores. El "hablante" es el interlocutor que pronuncia la falacia contra su "oponente", creando una versión distorsionada (falaz) de su posición real para luego refutarla. Las flechas representan el flujo de razonamiento en la argumentación. Algunos componentes son obligatorios (línea continua) y otros son opcionales (línea discontinua). La representación falaz de la posición del oponente y su refutación por el hablante son los dos componentes imprescindibles para la realización de la falacia, mientras que la posición real del oponente y la expresión de su supuesta refutación pueden entenderse de forma implícita o estar ausentes. La refutación siempre se marca con léxico explícito, aunque en algunos casos puede sustituirse o reforzarse por una contraposición del hablante contrastada con la posición falaz del oponente mediante léxico de contraste o por yuxtaposición de oraciones.

6.6 Reconocimiento de la falacia

Tal como se comentó anteriormente en la Sección 5.4, se ha planteado la identificación de la falacia como un proceso de reconocimiento de patrones a dos niveles: 1) reconocimiento de las clases de proposición típicas de la falacia y 2) reconocimiento del patrón composicional de estas proposiciones en un segmento delimitado de texto. Aplicando lo que el análisis de los datos ha revelado sobre estas dos tareas permite proponer un método o estrategia para la detección de la falacia del hombre de paja que se puede resumir como sigue, donde el "hablante" es el interlocutor que pronuncia la falacia contra su "oponente" en el discurso:

1) Identificar proposiciones falaces a partir de la presencia de los siguientes rasgos:
   - n-gramas con valores relativamente bajos de polaridad en el dominio tratado.
   - Nombres o pronombres correspondientes al oponente u otras personas relacionadas.
   - Verbos en 3ª (o 2ª) persona de las clases indicadas en (6.3) (que asocian ideas con oponente).
   - Función de sujeto del nombre/pronombre para el verbo.
   - Proximidad sintáctica de estos elementos.

2) Identificar refutaciones a partir de los siguientes rasgos:
   - Léxico clave de refutación (6.2).
   - Pronombres correspondientes al hablante (6.3).
   - Pronombres impersonales (6.2).
   - Proximidad sintáctica de estos elementos (pronombre justo antes de léxico de refutación).

3) Identificar proposiciones expresando la posición del hablante a partir de los siguientes rasgos:
   - Pronombres correspondientes al hablante.
   - n-gramas con valores relativamente neutros o positivos de polaridad en este dominio.
   - Verbos en 1ª persona de las clases indicadas en (6.3) (que asocian ideas con hablante).
   - Función de sujeto del pronombre para el verbo.
   - Proximidad sintáctica de estos elementos.

4) Detectar la composición de los 3 elementos anteriores (6.4):
   - Refutación dentro de proposición falaz o en frase justo anterior o posterior.
   - Una o más proposiciones de posición del hablante justo delante de refutación o en frase justo anterior o posterior.

Esta estrategia podría dar una primera aproximación a la detección de la falacia, aunque el grado de acierto probablemente variaría mucho en función de la complejidad sintáctica y discursiva del texto y de cuánto análisis semántico sea necesario para detectar el contenido inaceptable de las proposiciones falaces. Para una aplicación real también sería imprescindible emplear un corpus mucho más extenso y que haya sido anotado siguiendo un protocolo detallado y riguroso de anotación.

7 CONCLUSIONES

Con este trabajo se ha hecho el primer intento, que se conozca, de abordar el problema de minar una falacia argumental en un corpus lingüístico. La detección automática de falacias en la argumentación es un problema de gran relevancia hoy en día en el contexto de la lucha contra la deshonestidad en la política, la diseminación de bulos y noticas falsas, y la manipulación de la opinión pública, para citar algunos ejemplos. El estudio se ha centrado en la denominada *falacia del hombre de paja*, una clase de falacia común y muy prevalente en el discurso político.

Partiendo del planteamiento y la metodología del campo más afín, la minería de argumentos, y del estudio teórico y empírico de la falacia del hombre de paja, se ha formulado un modelo conceptual de esta clase de falacia, se ha diseñado una metodología para su identificación manual y anotación en un corpus lingüístico real, y se han propuesto rasgos de entrenamiento y una estrategia para su detección automática mediante algoritmos de clasificación.

Las principales aportaciones de este trabajo son:

- *Un modelo conceptual de la falacia del hombre de paja basado en su estudio teórico y el análisis de 85 ejemplos de la falacia extraídos de un corpus lingüístico real*. La falacia puede modelarse básicamente como:

    - *Una proposición falaz de contenido exagerado, inverosímil, absurdo, extremo, deplorable, caricaturizado, ridículo, sobregeneralizado o sobresimplificado* pronunciada por uno de los interlocutores en un discurso (el hablante) contra el otro (el oponente).

    - *La asociación de la proposición falaz con el oponente* del hablante mediante elementos lingüísticos específicos (generalmente nombres/pronombres en función de sujeto de ciertas clases de verbos).

    - Un claro *rechazo o refutación* de dicha proposición falaz por parte del hablante mediante el uso de léxico clave específico (generalmente léxico de negación con o sin pronombre personal/impersonal + verbo y/o adjetivos calificadores negativos), y/o mediante construcciones contrastivas con su propia posición.

- *Criterios y metodología para la identificación y anotación manual de la falacia en un corpus*. Los criterios básicos establecidos para la identificación manual de la falacia son:

- Se considera que una falacia está contenida en una única respuesta de un único interlocutor.
- No se contemplan casos en que una falacia abarque simultáneamente afirmaciones de ambos hablantes.
- Se presupone que cualquier afirmación inverosímil, deplorable, etc. asociada a un oponente es una distorsión de su posición real.
- La asociación de la proposición falaz y rechazada con el oponente no ha de ser siempre explícita, sino que puede entenderse de forma implícita.
- Los nombres y pronombres empleados para referirse al hablante y al oponente no siempre son los propios de esas personas.
- Las proposiciones que representan ideas y posiciones propias del hablante a menudo se anteponen o se posponen a la proposición falaz a modo de contraste y contraposición, que en algunos casos hace la función de refutación.

- *Análisis y anotación de 85 ejemplos de la falacia* en un corpus lingüístico real.
- *Identificación de patrones lingüísticos en el corpus* de utilidad para entrenar un algoritmo de detección automática de la falacia. Los patrones comprenden:
  - Léxico clave marcando la refutación de una proposición falaz.
  - Léxico clave marcando la asociación de proposiciones con interlocutores.
  - Secuencias sintácticas de los diferentes componentes de la falacia (proposiciones y léxico clave).
- *Propuesta de rasgos de entrenamiento* y una *estrategia de detección automática* de la falacia consistente en:

1) Identificar proposiciones falaces asociadas al oponente a partir de la presencia de n-gramas de polaridad negativa y su proximidad sintáctica a nombres y pronombres con función de sujeto de ciertas clases de verbos.

2) Identificar refutaciones a partir de la presencia de léxico clave y pronombres con proximidad sintáctica.

3) Identificar proposiciones expresando la posición del hablante a partir de la presencia de n-gramas de polaridad neutra o positiva y su proximidad sintáctica a pronombres con función de sujeto de ciertas clases de verbos.

4) Detectar patrones secuenciales de los 3 elementos anteriores.

Sería interesante una futura continuación de este trabajo para evaluar la metodología propuesta aplicándola al desarrollo un sistema de reconocimiento automático. Previsiblemente, la detección del carácter inaceptable de las proposiciones falaces sería un factor crítico para su buen funcionamiento, por lo que sería recomendable un análisis en mayor profundidad del problema de evaluar su contenido semántico mediante la asignación de polaridades a n-gramas o incluso sintagmas basada en conocimiento del dominio del texto tratado.

REFERENCIAS

BADERTSCHER (2017), Using Corpus Analysis Tools to Compare the Rhetorical Styles of Hillary Clinton and Donald Trump in the 2016 Presidential Debates, trabajo fin de máster, Universitat de Barcelona.

CARLSON, L., MARCU, D. y OKUROWSKI, M.E. (2002), RST Discourse Treebank, Technical Report LDC2002T07, Linguistic Data Consortium, Philadelphia.

CHIASMA (2016), 5 Logical Fallacies from the Second Clinton Trump Debate, en *Medium* (blog). <https://medium.com/@Chiasma/5-logical-fallacies-from-the-second-clinton-trump-debate-4080336314e1>

Corpus of 2016 American Presidential Debate responses. <https://drive.google.com/open?id=0B76VPd6ZCQ-BYjhtZmlwM09yblU>

HABERNAL, I., GUREVYCH, I. (2016), Argumentation Mining in User-Generated Web Discourse, *Computational Linguistics*, 43(1): 125-179.

LIPPI, M., TORRONI, P., (2016), Argumentation mining: State of the art and emerging trends, *ACM Transactions on Internet Technology*, 16(2): 1–25.
<http://lia.disi.unibo.it/~ml/publications/ACMTOIT2015.pdf>

LIPPI, M., TORRONI, P. (2015), Argument Mining: a Machine Learning Perspective, *The 2015 International Workshop on Theory and Applications of Formal Argument*, Buenos Aires.
<http://homepages.abdn.ac.uk/n.oren/pages/TAFA-15/TAFA-15_submission_18.pdf>

HANSEN, H. (2015), Fallacies, en *The Stanford Encyclopedia of Philosophy*, Stanford University. <https://plato.stanford.edu/entries/fallacies/>

KHARE, A. (2016), Latticework of Mental Models: Straw Man Fallacy, en *Safal Niveshak*. https://www.safalniveshak.com/latticework-mental-models-straw-man-fallacy/

ROITMAN, M. (2017), Negation and straw man fallacy in French election debates 1974-2012, en *The Pragmatics of Negation*, Malin Roitman Ed., John Benjamin Publishing Company, Amsterdam.

SANJUL, S. (2018), Ocho falcacias lógicas con los que los políticos hackean tu mente, en Tribus Ocultas (web), <http://www.lasexta.com/tribus-ocultas/libros/ocho-falacias-logicas-que-politicos-hackean-mente_20170712599caa470cf2e2ea355e6011.html>

TOULMIN, S. (1958), The Uses of Argument, Cambridge University Press.

University of Minnesota (2016), Informative and Persuasive Speaking, en *Communication in the Real World: An Introduction to Communication Studies*, Ch. 11.3, University of Minnesota (autores no especificados por editor). <http://open.lib.umn.edu/communication/chapter/11-3-persuasive-reasoning-and-fallacies/>

VAN ONSELER, G. (2012), The straw man fallacy, en www.inside-politics.org (blog), <https://inside-politics.org/2012/02/20/the-straw-man-fallacy/>

WALTON, D. (1996), The straw man fallacy, en *Logic and Argumentation*, Johan van Bentham, Frans H. van Eemeren, Rob Grootendorst y Frank Veltman (eds), Amsterdam: 115-128.

Wikipedia (2018), List of Fallacies, Wikipedia. <https://en.wikipedia.org/wiki/List_of_fallacies>

ZURLONI V., ANOLLI L. (2013) Fallacies as Argumentative Devices in Political Debates, en: *Multimodal Communication in Political Speech. Shaping Minds and Social Action. Lecture Notes in Computer Science*, Poggi I., D'Errico F., Vincze L., Vinciarelli A. (eds), vol 7688. Springer, Berlin, Heidelberg: 245-257.

ANEXO 1: LÉXICO CLAVE Y PATRONES SINTÁCTICOS

Hillary Clinton						
índice	**patrón sintáctico**	**pronombre hablante**	**pronombre oponente**	**léxico asociación**	**pronombre refutación**	**refutación**
HRC1.12	1**RD**	-	-	-	us	let's not
HRC1.42	1**RD**	-	-	-	I	don't think
HRC1.63	1**RD**	-	-	-	we	are not going to
HRC3.6	1**RD**	-	-	-	I	do not think
HRC3.8	1**RD**	-	-	-	I	don't want
HRC3.8.2	1**RD**	-	-	-	I	don't want
HRC3.34	1**RD**	-	-	-	I	will not
HRC3.8.3	1**R**3**D**	-	Donald	has talked about	I	don't want
HRC2.50.5	1**R**3**D**	I	you	want	-	doesn't mean
HRC2.22	1**R**3**D**	us	Donald	is important for, has said	it	not
HRC3.75	1P.1**RD**	I	-	am not going to	I	am not going to
HRC1.55	1P.1**RD**	we	-	have got to	we	cannot
HRC3.7.2	**RD**.1P	I	-	will stand up for	the government	has no business
HRC3.80	1**RD**.1P	I	-	want	I	will not
HRC1.2	1P**RD**	we	-	have to	-	not
HRC1.2.2	1P**RD**	I	-	want	-	not
HRC1.3	1P**RD**	we	-	need	-	not
HRC1.56	1P**RD**	I	-	think	-	not
HRC2.28	1P**RD**	we	-	are going to	-	nobody, no
HRC2.43.3	1P**RD**	we	-	ran	-	not
HRC2.50	1P**RD**	I	-	want	-	not
HRC2.50.4	1P**RD**	I	-	want	-	doesn't
HRC3.1	1P**RD**	I	-	feel	-	not
HRC3.1.3	1P**RD**	I	-	support, want	-	doesn't
HRC2.1	1P+1P**RD**	I,we	-	think, have set forth	-	not
HRC1.7	1P+1P**RD**	I,we	-	have, should, am determined	-	never
HRC1.42.2	1P+1P**RD**	we	-	need	-	not
HRC1.24	1P.1P**RD**	I,we	-	think, need, said	-	not
HRC3.64	1P.1P1**RD**	I,we	-	have said, need, want, have to	we	don't let
HRC2.50.2	1P**RD**.1P**RD**.3RT	I	Donald	think, would want, doesn't agree	Donald	wrong, reverse, out of, doesn't agree with
HRC1.43	1P**RD**.3**D**	we	he	doesn't want, know, does, paid, had to, was trying to, didn't pay	there	is no
HRC2.27	1P**RD**.1P	I	him, his	want, said	he	should not
HRC2.54	3**D**.1P**RD**	I	Donald Trump	is buying, would have	I, we	fought against, don't

HRC1.68.2	1P**R**3**D**	they	Donald's	need, argument is	-	not
HRC2.43	1P**R**3**D**	I,my	his,him,he	said, argument is, has run, inciting	-	is not
HRC3.10	1P**R**3**D**	I	Donald	exploit	-	not
HRC3.15	1P**R**3**D**	I	Donald Trump, he	think, admit, condemn, rejects, encouraged	-	will not have
HRC1.44	3D.1**R**3**D**	I	he	think, is not that enthusiastic about, trying to	-	think probably he's not
HRC1.50	3**DR**	-	your	refused	-	do..not
HRC2.39	1P.1P.3**D.R**	I	Donald, he	hope, do think, says	-	no, doesn't
HRC1.3.3	3**D**.1**R**	-	he, his	believes	I	don't buy
HRC1.3.2	1**RD**.T**R**	I	-	call it	that	is not
HRC3.36	1P**RD**.T**R**	I	-	am proposing	that	is not
HRC2.20	1P+1**RD**.T**R**	us	-	let	that	let's not, is not
HRC3.37-38	3**D**.T**R**	-	Donald, Trump, he	has, goes around with	that	is not going to work
HRC1.71	3**D**.T**R**	-	Donald, he,(I)	saying, said	that	is not good
HRC2.27.2	3**D**T**R**	-	Donald	paid	that	is wrong
HRC3.48	T3**D**.T**R**	-	he, his	incites, applauds	this	is not
HRC2.50.3	3**D**.1PT**R**	I	Donald, he	has put forth, would consider, has suggested, think	that	would be..mistake + backwards
HRC1.28-29	1P.1P+3**D**.T**R**	we, our, mine	yours	have..plans	that	can't be
HRC3.34.2	1P.3**D**	I	he, his	am going to, will	-	-
HRC3.62	1P.1P.1P.3**D**	we	you	have accepted, may not have liked, are whining	-	-
HRC2.43.2	3**D**.1P	I	he	has said, has never apologized for, do think,has said, am proud of, ran	-	-
HRC1.68	3**D**C1P	we	Donald	need	-	when
HRC3.32	TP.C3**D**	this	Donald's, his	is a plan, plan, plan is	-	by contrast

Tabla 2: Léxico clave y patrones sintácticos en las falacias del hombre de paja pronunciadas por Hillary Clinton.

Donald Trump						
índice	**patrón sintáctico**	**pronombre hablante**	**pronombre oponente**	**léxico asociación**	**pronombre refutación**	**refutación**
DT3.118	1**RD**	we	-	-	-	are not going to
DT3.128	1**R**3**D**	we	her	-	-	cannot
DT3.1	1**R**3**D**	I	we	-	-	don't think
DT1.108.2	1P.1**RD**.1RD	I, we	-	want	-	cannot
DT3.10.2	3D.3D.1P.1P.1P.3D.1P.1**RD**	We	Hillary Clinton, Barack Obama	caused by, have to, need	-	cannot
DT1.36	**R**3**D**	-	you	are telling	-	no
DT2.87.3	1P1**R**3**D**	I	she	will have, is doing	-	am not
DT1.104	3**DR**.3R.3R	-	they	made	-	horrible, should have
DT2.17.5	3**DR**.3D	-	her, you	to say, did	-	should be
DT1.35	3**D**.1**R**	I	her, she	is telling, tells	-	don't think
DT1.59	3**D**.1**R**	we	Secretary Clinton	doesn't want	-	need
DT2.17	3D.3**D**.1**R**	I	Hillary, she, you	said, did, think	-	don't think
DT1.38	3D.3**D**.1**R**1P.3D.3D	everybody	Secretary Clinton, they	won't allow, can't work out, agrees, can't get together, can't sit them around	-	should be
DT2.17.3	3D.3D.3**D**.1**R**	I	she, her	said, is saying, will be honest	-	am so disappointed
DT2.49	1R.3D.3D.3D.3**D**.1**R**	I	she, her	think,said,got caught in,is blaming	-	so ridiculous, haven't
DT2.13.2	3D.3D+3**D**.1**R**.1P	me, I	Hillary Clinton, her, she	attacked, represented, has been seen laughing, am, apologize for	-	don't
DT2.88.3	3D.1P.3**D**.1**R**	we, I	Hillary Clinton, we	wants, will bring, are	-	have to
DT1.40	T**R**D	-	-	-	there	is nothing crazy about not
DT3.116	T**R**3**D**.1**R**D	we	she	will	it, we	is so ridiculous, should have never
DT1.104.3	3**D**.T**R**	-	they	didn't make	this	is one of the worst
DT1.107-108	3**D**.T**R**	-	Hillary, she	will tell	it	is a big problem
DT2.48	3**D**.T**R**	-	we	letting, shouldn't be	it	shouldn't
DT1.104.2	3D.3**D**.T**R**	-	Secretary Kerry, you	didn't do, didn't add	that	turned out to be wrong
DT2.13	3R.1P+3D.3**D**.T**R**D.3D	mine	his, he, Bill Clinton	are, was, has done, was	there	has never been
DT2.48.2	1P.3D.3**D**.T**R**3D.3DR	I	she	understand, doesn't, wants, is doing, has got, honestly, can tell	it	should never be, is a horrible thing, bad
DT2.13.4	3**D**T**R**+1**R**	I	Hillary, she	brings up a point, talks about, think	it	is disgraceful, should be ashamed
DT3.8	3**D**.T**R**1R**3D**.**TR**	me	Hillary, she	can say,is not OK with,is saying,is going,has been,is not	it, that	is not OK, is not acceptable
DT2.87	1P.1P+1P.1P.3**D**.1P	I, me	Hillary Clinton	think,is..under siege by, important to	-	-

Tabla 3: Léxico clave y patrones sintácticos en las falacias del hombre de paja pronunciadas por Donald Trump.

UNIVERSITAT DE BARCELONA

Coordinació d'Estudis
Facultat de Filologia

Gran Via
de les Corts Catalanes, 585
08007 Barcelona

Tel. +34 934 035 594
fil-coord@ub.edu
www.ub.edu

Declaració d'autoria

Amb aquest escrit declaro que sóc l'autor/autora original d'aquest treball i que no he emprat per a la seva elaboració cap altra font, incloses fonts d'Internet i altres mitjans electrònics, a part de les indicades. En el treball he assenyalat com a tals totes les citacions, literals o de contingut, que procedeixen d'altres obres. Tinc coneixement que d'altra manera, i segons el que s'indica a l'article 18, del capítol 5 de les Normes reguladores de l'avaluació i de la qualificació dels aprenentatges de la UB, l'avaluació comporta la qualificació de "Suspens".

Barcelona, a 20 de junio de 2018

Signatura: